



AF E ZMW

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

ELIOT M. CASE et al.

Serial No.: 09/872,680

Filed: June 1, 2001

For: METHOD AND APPARATUS FOR RECORDING PROSODY FOR FULL  
CONCATENATED SPEECH

Group Art Unit: 2655

Examiner: Huyen X. Vo

Attorney Docket No.: 1809 / USW 0613 PUS

APPEAL BRIEF UNDER 37 C.F.R. § 41.37

Mail Stop Appeal Brief - Patents  
Commissioner for Patents  
U.S. Patent & Trademark Office  
P.O. Box 1450  
Alexandria, VA 22313-1450

Sir:

This is an Appeal Brief from the final rejection of claims 1-18 of the Office  
Action mailed on October 29, 2004 for the above-identified patent application.

I. REAL PARTY IN INTEREST

The real party in interest is Qwest Communications International Inc.  
("Assignee"), a corporation organized and existing under the laws of the state of Delaware,  
and having a place of business at 1801 California Street, 9<sup>th</sup> Floor, Denver, Colorado 80202.

05/05/2005 MAHMED1 00000053 09872680

01 FC:1402 500.00 OP

05/05/2005 MAHMED1 00000053 09872680

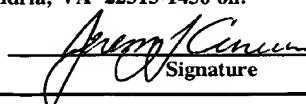
02 FC:1251 120.00 OP

**CERTIFICATE OF MAILING UNDER 37 C.F.R. § 1.8**

I hereby certify that this paper, including all enclosures referred to herein, is being deposited with the United States Postal Service as first-class mail, postage pre-paid, in an envelope addressed to: Mail Stop Appeal Brief - Patents, Commissioner for Patents, U.S. Patent & Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450 on:

May 2, 2005  
Date of Deposit

Jeremy J. Curcuri  
Name of Person Signing

  
Signature

## **II. RELATED APPEALS AND INTERFERENCES**

There are no appeals or interferences known to the Appellant, the Appellant's legal representative, or the Assignee which will directly affect or be directly affected by or have a bearing on the Board's decision in the pending appeal.

## **III. STATUS OF CLAIMS**

Claims 1-18 are pending in this application. Claims 1-18 have been rejected and are the subject of this appeal.

## **IV. STATUS OF AMENDMENTS**

No amendment after final rejection has been filed.

## **V. SUMMARY OF CLAIMED SUBJECT MATTER**

The invention relates to a system and method for converting text to voice. Page 1, lines 5-6. As shown in Figure 1, a digital voice library 12 is an asset database that includes human voice recordings of syllables, words, phrases, and sentences in a significant number of voice inflections. In a converting operation, systems and methods perform analysis of incoming text 14, and access digital voice library 12 via look-up logic 16 for voice recordings with the desired prosody or inflection, and pronunciation. Sentence construction algorithms 18 are employed to concatenate together spoken sentences or voice output 20 of the text input. Page 8, lines 6-18. Figure 2 illustrates the architecture and flow of a preferred text to voice conversion system and method.

The invention comprehends a method of making a digital voice library 12 utilized for converting text to concatenated voice in accordance with a set of playback rules 98. The digital voice library 12 includes a plurality of speech items and a corresponding plurality of voice recordings. Each speech item corresponds to at least one available voice recording.

Multiple voice recordings that correspond to a single speech item represent various inflections of that single speech item. Page 1, line 20-Page 2, line 1.

The method, as best illustrated in Figure 26, comprises establishing a vocal sequence (block 102). A voice talent is recorded uttering the vocal sequence (block 104). A complex tone that reflects a particular inflection required for a particular voice recording of a particular speech item is generated (block 106). The complex tone is composed of portions of the recording of the voice talent uttering the vocal sequence. The method further comprises recording the voice talent reciting the particular speech item to make the particular voice recording. The voice talent uses the complex tone as a guide to allow the voice talent to recite a particular speech item in accordance with the particular inflection (block 108). Page 2, lines 1-9. Page 38, line 7-Page 39, line 10. Page 40, lines 14-30.

According to the invention, the method of making a digital voice library may be used to make voice recordings for any speech items including phonemes, syllables, words, phrases, and/or sentences. In addition, it is appreciated that establishing the vocal sequence and recording the voice talent may include uttering the vocal sequence by speaking, humming, or singing, or any other technique. Page 2, lines 10-20, Page 40, line 30-Page 41, line 5.

In accordance with the invention, the complex tone is a complex wave form recorded in the voice talent's own voice, using the complex tone as a guide makes it easier for the voice talent to synchronize with the complex tone because the complex tone is made up of the voice talent of the actual voice. Page 3, lines 7-13.

The invention also comprehends a digital voice library wherein the voice recordings are made using the comprehended methods. Page 2, line 21-Page 3, line 6.

## **VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL**

Claims 1-18 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over Gasper et al. (U.S. Patent No. 5,278,943) in view of Tubman et al. (U.S. Patent No. 5,820,384).

## **VII. ARGUMENT**

### **A. Claims 1-18 Are Patentable Over Gasper In View Of Tubman**

Claims 1-18 stand finally rejected under 35 U.S.C. § 103(a) as being unpatentable over Gasper et al. (US Patent No. 5,278,943) in view of Tubman et al. (US Patent No. 5,820,384).

Applicants' invention generally comprehends a method and apparatus for recording prosody for fully concatenated speech wherein a digital voice library and a method of making a digital voice library for use in text to concatenated voice applications are disclosed.

Regarding the rejection of claims 1 and 10, a digital voice library and a method of making a digital voice library used for converting text to concatenated voice in accordance with a set of playback rules are recited. The digital voice library includes a plurality of speech items and a corresponding plurality of voice recordings wherein each speech item corresponds to at least one available voice recording. The multiple voice recordings that correspond to a single speech item represent various different inflections of that single speech item.

The method comprises establishing a vocal sequence and then recording the voice talent uttering the vocal sequence. A complex tone is generated that reflects a particular inflection required for a particular voice recording of a particular speech item. The complex tone is composed of portions of the recording of the voice talent uttering the vocal sequence.

The voice talent is recorded reciting the speech item to make the particular voice recording. The voice talent uses the complex tone as a guide to allow the voice talent to recite the particular speech item in accordance with the particular inflection.

Specifically, the complex tone acts as a guide from which the vocal talent is to follow when reciting the vocal sequence to provide a particular inflection for a particular voice recording. The digital voice library is generated from speech items representing various inflections recorded as recited by the voice talent specifically using the complex tone composed of the voice talent's own utterances as a guide.

However, in Gasper pre-recorded speech samples retrieved from a library are processed to add inflection and other auditory effects to create animated or artificial voices. Gasper merely describes a voice animation system whereby pre-recorded speech samples are divided into basic segments for use in a text to speech synthesizer to artificially synthesize speech. The voice talent does not recite vocal sequences with the proper inflection while using a complex tone composed of the voice talent's own utterances as a guide, but rather, it is the pre-recorded samples that are processed after being recorded to add inflection and other auditory effects to create animated or artificial voices according to a prosody rule set.

In Applicant's invention, a complex tone is generated "that reflects a particular inflection required for a particular voice recording of a particular speech item" and further "recording the voice talent reciting the particular speech item to make the particular voice recording, the voice talent using the complex tone as a guide to allow the voice talent to recite the particular speech item in accordance with the particular inflection." Thus, in Gasper, the speech animation and inflections are synthesized in a second stage after the segments are retrieved from the library and speech output is then processed from the pre-existing segments.

The Examiner recognizes that Gasper fails to specifically disclose a method for generating a complex tone that reflects a particular inflection required for a particular voice recording of a particular speech item. Also, Gasper fails to describe or suggest the complex tone being composed of portions of the recording of the voice talent uttering the vocal sequence and for recording the voice talent reciting the particular speech item, the voice talent using the complex tone as a guide to allow the voice talent to recite the particular speech item in accordance with the particular inflection.

The Examiner relies on Gasper in view of Tubman to make the rejection. Applicant contends that Tubman fails to overcome the deficiencies of Gasper, and further, that there is no motivation to combine the teachings of the references in such a way to achieve the claimed invention.

Gasper fails to recognize or suggest a need for generating a complex tone having a particular inflection needed for a particular recording from portions of the recording of the voice talent or for using the complex tone as a guide by the voice talent to recite specific utterances having specific inflections for making recordings necessary to generate the digital voice library. Thus, Gasper fails to provide the required motivation to combine the references.

Tubman merely describes a recording method and system for providing acoustical prompts for Karaoke participants. The Tubman method employs a listen-sing-along procedure effected via the interaction of the spoken instructor-promptings and the Karaoke participant. There is no suggestion that any of the teachings of Tubman would be useful in a method of making a digital voice library used for converting text to concatenated voice in accordance with a set of playback rules. Additionally, Tubman does not suggest modifying the system taught by Gasper to achieve Applicant's invention.

As such, there is no suggestion or motivation to combine the voice animation system of Gasper and the Karaoke system described in Tubman to achieve the claimed invention. After all, Tubman only describes a recording method and system of prompting a message accordingly with the melody of a vocal line, but only in the very limited application of enabling a Karaoke participant to generate his/her own renditions of vocal in accompanying relationship with music.

Further, Tubman is believed to be non-analogous art. The invention relates to a system and method for converting text to voice. The invention addresses specific problems associated with making digital voice libraries. Tubman is not in the same field of endeavor. Tubman relates to a karaoke sing-along method and system using acoustical prompting rather than visual prompting. In contrast, the invention relates to systems and methods for converting text to voice. Further, Tubman does not logically commend itself to an inventor's attention when addressing problems associated with digital voice libraries.

For the reasons given above, Tubman is believed to be non-analogous art. Further, any combination of Gasper and Tubman is believed to be deficient, and there is no motivation to combine these references to achieve the claimed invention.

The remaining claims, namely, claims 2-9 and 11-18 are dependent claims and are also believed to be patentable.

In the final action, the Examiner states that "the advantage of combining the teaching of Tubman et al. and Gasper et al. is to assist music listeners to practice songs by singing along with the karaoke system." This advantage bears no relation to the claimed invention. The claimed invention is a digital voice library and a method of making a digital voice library that recites specific features in combination involving the use of a complex tone composed of portions of the recording of the voice talent uttering the vocal sequence as a guide

to allow the voice talent to recite the particular speech item in accordance with the particular inflection. Thus, in response to the Examiner's response to arguments, Applicants maintain that there is no motivation to combine the references to achieve the claimed invention. Allowing music listeners to practice songs by singing along with the karaoke system fails to suggest any motivation to use features from the sing along karaoke system in a digital voice library or method of making a digital voice library for use in text to concatenated voice applications.

The fee of \$500 as applicable under the provisions of 37 C.F.R. § 41.20(b)(2) is enclosed. Also included is the fee of \$120 for a one month extension of time. Please charge any additional fee or credit any overpayment in connection with this filing to our Deposit Account No. 02-3978.

Respectfully submitted,  
ELIOT M. CASE et al.

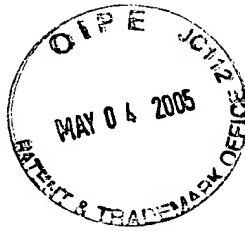
By: Jeremy J. Carcuri  
Jeremy J. Carcuri  
Registration No. 42,454  
Attorney for Applicant

Date: May 2, 2005

**BROOKS KUSHMAN P.C.**  
1000 Town Center, 22nd Floor  
Southfield, MI 48075-1238  
Phone: 248-358-4400  
Fax: 248-358-3351

Enclosure - Appendices





### **VIII. CLAIMS APPENDIX**

1. A method of making a digital voice library utilized for converting text to concatenated voice in accordance with a set of playback rules, the digital voice library including a plurality of speech items and a corresponding plurality of voice recordings wherein each speech item corresponds to at least one available voice recording, wherein multiple voice recordings that correspond to a single speech item represent various inflections of that single speech item, the method comprising:

establishing a vocal sequence;

recording a voice talent uttering the vocal sequence;

generating a complex tone that reflects a particular inflection required for a particular voice recording of a particular speech item, the complex tone being composed of portions of the recording of the voice talent uttering the vocal sequence; and

recording the voice talent reciting the particular speech item to make the particular voice recording, the voice talent using the complex tone as a guide to allow the voice talent to recite the particular speech item in accordance with the particular inflection.

2. The method of claim 1 wherein establishing the vocal sequence and recording the voice talent further comprise:

establishing the vocal sequence as a sequence of words; and

recording the voice talent speaking the sequence of words.

3. The method of claim 1 wherein establishing the vocal sequence and recording the voice talent further comprise:

establishing the vocal sequence as a sequence of tones; and

recording the voice talent humming the sequence of tones.

4. The method of claim 1 wherein establishing the vocal sequence and recording the voice talent further comprise:

establishing the vocal sequence as a sequence of words; and  
recording the voice talent singing the sequence of words.

5. The method of claim 1 wherein the particular speech item is a phoneme.

6. The method of claim 1 wherein the particular speech item is a syllable.

7. The method of claim 1 wherein the particular speech item is a word.

8. The method of claim 1 wherein the particular speech item is a phrase.

9. The method of claim 1 wherein the particular speech item is a sentence.

10. A digital voice library utilized for converting text to concatenated voice in accordance with a set of playback rules, the digital voice library including a plurality of speech items and a corresponding plurality of voice recordings wherein each speech item corresponds to at least one available voice recording, wherein multiple voice recordings that correspond to a single speech item represent various inflections of that single speech item, the digital voice library further comprising a particular voice recording of a particular speech item, the particular voice recording requiring a particular inflection and being made by:

establishing a vocal sequence;

recording a voice talent uttering the vocal sequence;

generating a complex tone that reflects the particular inflection required for the particular voice recording of the particular speech item, the complex tone being composed of portions of the recording of the voice talent uttering the vocal sequence; and

recording the voice talent reciting the particular speech item to make the particular voice recording, the voice talent using the complex tone as a guide to allow the voice talent to recite the particular speech item in accordance with the particular inflection.

11. The digital voice library of claim 10 wherein establishing the vocal sequence and recording the voice talent further comprise:

establishing the vocal sequence as a sequence of words; and  
recording the voice talent speaking the sequence of words.

12. The digital voice library of claim 10 wherein establishing the vocal sequence and recording the voice talent further comprise:

establishing the vocal sequence as a sequence of tones; and  
recording the voice talent humming the sequence of tones.

13. The digital voice library of claim 10 wherein establishing the vocal sequence and recording the voice talent further comprise:

establishing the vocal sequence as a sequence of words; and  
recording the voice talent singing the sequence of words.

14. The digital voice library of claim 10 wherein the particular speech item is a phoneme.

15. The digital voice library of claim 10 wherein the particular speech item is a syllable.

16. The digital voice library of claim 10 wherein the particular speech item is a word.

17. The digital voice library of claim 10 wherein the particular speech item is a phrase.
18. The digital voice library of claim 10 wherein the particular speech item is a sentence.